STATISTICAL RESEARCH REPORT No 2
Institute of Mathematics March 1977
University of Oslo

EMPIRICAL LINEAR CREDIBILITY ESTIMATORS

by

Ragnar Norberg

1. Introduction. The model and the problem

1 A. Let Θ assume its values randomly in some space Ω according to a probability measure U. The value of Θ is unobservable, but with Θ is associated an observable random vector $\underline{X} = (X_1,\dots,X_m)'$, whose conditional distribution function, given $\Theta=\theta$, is $F(\circ|\theta)$. By means of $\underline{X}$ we shall estimate a scalar valued function of Θ, say $g(\Theta)$. The goodness of an estimator $\hat{g}$ is measured by the expected squared error

$$E\{g(\Theta)-\hat{g}\}^2 . \tag{1.1}$$

This situation fits the Bayesian scheme of estimation, with prior measure U, (class of) kernel distributions

$$\mathcal{F} = \{F(\circ|\theta) \; ; \; \theta \in \Omega\}$$

and quadratic loss function. We shall, however, admit as estimators only functions of the form

$$\hat{g} = a_o + \sum_{i=1}^{m} a_i X_i . \tag{1.2}$$

Thus, so far our problem is that of multidimensional linear credibility estimation, treated by Jewell (1973). Accordingly we will speak of the set of coefficients $\alpha_o,\dots,\alpha_m$ minimizing (1.1) as the <u>credibility coefficients</u> and the optimal linear formula

$$\tilde{g} = \alpha_o + \sum_{i=1}^{m} \alpha_i X_i \tag{1.3}$$

as the <u>linear credibility estimator.</u> This terminology is borrowed from the theory of individual experience rating, where Θ is a risk parameter, measuring the accident proneness of the risk, and $\underline{X}$ is the statistical information generated by the risk, typically including its claims records in earlier insurance periods. On basis of $\underline{X}$ the expected amount of claims in the next insurance period, $g(\Theta)$, is to be estimated. The simple derivation of the credibility formula is demonstrated in subsection 2 A.

1 B. Of particular interest is the linear regression case when

$$E(\underline{\underline{X}} | \Theta = \theta) = \underline{\underline{Y}}\, \underline{\underline{B}}(\theta)$$

for some known, nonrandom "design matrix" $\underline{\underline{Y}}$ of order $m \times q$, say, and vector valued function $\underline{\underline{B}}$ of order $q \times 1$. In this situation we typically want to estimate $g(\Theta) = \underline{\underline{y}}'\underline{\underline{B}}(\Theta)$ for some fixed q-vector $\underline{\underline{y}}$, the class of admissible estimators and the measure of goodness still being as in subsection 1 A. This is the problem of estimation by a non-homogeneous credibility formula, investigated by Hachemeister (1975) and in a more general setting by Taylor (1975). In subsection 2 B the solution will be derived from that of the general problem presented in subsection 1 B. The present model includes random regression coefficient models of the type studied first by Wald (1947) and later by many others. For a review, see Swamy (1971).

1 C. The credibility coefficients depend on the moments up to second order of the variables $g(\Theta), X_1, \ldots, X_m$. In many situations reasonable assumptions will lead to a known family $\mathcal{F}$ of conditional distributions. However, in practical situations the measure U, and hence also the linear credibility formula (1.3), are unknown. To remedy this problematic state of affairs, Bühlmann and Straub (1970) proposed to replace the unknown moments involved in the credibility coefficients by unbiased estimates. These estimates will typically be derived from n independent vectors of observations $\underline{\underline{X}}_1, \ldots, \underline{\underline{X}}_n$ related to n preceding independent replications $\Theta_1, \ldots, \Theta_n$ of Θ. At this moment we need not specify further assumptions concerning the distribution of these observations. We recognize the problem of estimating $g(\Theta)$ by means of $\underline{\underline{X}}$ and $\underline{\underline{X}}_1, \ldots, \underline{\underline{X}}_n$, when U and possibly also $\mathcal{F}$ are unknown, as one essentially within the framework of empirical Bayes theory. This theory was founded by Robbins (1955 and 1964), and its main aspects and results are surveyed by Maritz

(1970). The purpose of this paper is to investigate the properties of Bühlmann-Straub type of estimators from the empirical Bayes point of view. Thus let $(\hat{a}_{no},\ldots,\hat{a}_{nm})$ be a function of $\underline{X}_1,\ldots,\underline{X}_n$ which converges in probability to the vector of credibility coefficients, $(\alpha_0,\ldots,\alpha_m)$, as $n \to \infty$. In section 3 it will be proved, under reasonable conditions, that the sequence of estimators

$$\hat{g}_n = \hat{a}_{no} + \sum_{i=1}^{m} \hat{a}_{ni} X_i$$

is asymptotically optimal in the sense that

$$\lim_{n\to\infty} E\{g(\Theta)-\hat{g}_n\}^2 = E\{g(\Theta)-\tilde{g}\}^2 \tag{1.4}$$

for all $(U,\mathfrak{F})$ such that the variances of the variables $g(\Theta)$, $X_1,\ldots,X_m$ exist, the expectation on the left side in (1.4) being with respect to Θ, $\underline{X}$, $\underline{X}_1,\ldots,\underline{X}_n$. In section 4 we first construct consistent estimators of the coefficients $\alpha_0,\ldots,\alpha_m$ in the linear regression case, and compare our results to those obtained by Martz and Krutchkoff (1969) and Wind (1973) for this special case. Secondly we consider a nested classification model.

2. Linear credibility estimators

2 A. In this section we recall how to find the credibility coefficients $\alpha_0,\ldots,\alpha_m$ which minimize (1.1) for a fixed pair $(U,\mathfrak{F})$. We denote by $\mathcal{M}_2$ the family of pairs $(U,\mathfrak{F})$ with the property that all the variables $g(\Theta)$, $X_1,\ldots,X_m$ have finite moments up to second order. In order that the expectation (1.1) be finite, and the problem of minimizing it thus well defined, we assume that $(U,\mathfrak{F}) \in \mathcal{M}_2$. By this assumption we may differentiate (1.1) with respect to the a_i's by reversing the order of expectation

and differentiation. This gives the following necessary conditions for $\alpha_o,\ldots,\alpha_m$ to minimize (1.1).

$$E\{g(\Theta)-\alpha_o-\sum_{i=1}^{m}\alpha_i X_i\}(-1) = 0 ,$$

$$E\{g(\Theta)-\alpha_o-\sum_{i=1}^{m}\alpha_i X_i\}(-X_k) = 0 , \quad k=1,\ldots,m .$$

Since (1.1) is a positive definite quadratic form in $a_o,\ldots,a_n$, these conditions are also sufficient. Multiplying the first equation by $E(X_k)$ before subtracting it from the k'th of the remaining equations, we arrive at

$$E\{g(\Theta)\} = \alpha_o+\sum_{i=1}^{m}\alpha_i EX_i , \tag{2.1}$$

$$\operatorname{cov}\{g(\Theta),X_k\} = \sum_{i=1}^{m}\alpha_i \operatorname{cov}(X_i,X_k) , \quad k=1,\ldots,m .$$

We have assumed that $(U,\mathcal{F}) \in \mathcal{M}_2$. This implies the existence and finiteness of the scalar

$$\gamma = E\{g(\Theta)\} ,$$

the vectors

$$\underline{\underline{M}}(\Theta) = E(\underline{\underline{X}}|\Theta) \quad , \quad \underline{\underline{\mu}} = E(\underline{\underline{X}}) = E\{\underline{\underline{M}}(\Theta)\} \tag{2.2}$$

$$\underline{\underline{\xi}} = [\operatorname{cov}\{g(\Theta),M_1(\Theta)\},\ldots,\operatorname{cov}\{g(\Theta),M_m(\Theta)\}]' , \tag{2.3}$$

and the matrices

$$\underline{\underline{\Delta}} = E\{\operatorname{Var}(\underline{\underline{X}}|\Theta)\} \quad \text{and} \quad \underline{\underline{\Gamma}} = \operatorname{Var}\{\underline{\underline{M}}(\Theta)\} ,$$

where Var denotes the covariance matrix operator.

Note that we have

$$\operatorname{Var}(\underline{\underline{X}}) = \underline{\underline{\Delta}} + \underline{\underline{\Gamma}}$$

and

$$\xi_i = \operatorname{cov}\{g(\Theta),X_i\} .$$

By means of these notations and relations we may write (2.1) as

$$\gamma = \alpha_0 + \underline{\underline{\zeta}}'\underline{\underline{\mu}}$$

$$\underline{\underline{\xi}} = (\underline{\underline{\Delta}} + \underline{\underline{\Gamma}})\underline{\underline{\zeta}} ,$$

where

$$\underline{\underline{\zeta}} = (\alpha_1,\ldots,\alpha_m)'.$$

The credibility coefficients are thus given by

$$\underline{\underline{\zeta}} = (\underline{\underline{\Delta}} + \underline{\underline{\Gamma}})^{-1}\underline{\underline{\xi}} \tag{2.4}$$

and

$$\alpha_0 = \gamma - \underline{\underline{\zeta}}'\underline{\underline{\mu}} .$$

Inserting these expressions in (1.3) we find that the linear credibility estimator is

$$\tilde{g} = \gamma + \underline{\underline{\zeta}}'(\underline{\underline{X}}-\underline{\underline{\mu}}) . \tag{2.5}$$

With this formula we estimate $g(\Theta)$ by its average value in the population adjusted according to the deviation of the observations $X_1,\ldots,X_m$ from their expected values. The credibility coefficients depend on $(U,\mathcal{F})$ through the first and second order moments

$$\underline{\underline{\tau}} = (\gamma,\ \underline{\underline{\mu}},\ \underline{\underline{\xi}},\ \underline{\underline{\Delta}} + \underline{\underline{\Gamma}}) .$$

For later use we introduce also the notation

$$\rho(\underline{\underline{\tau}}) = E\{g(\Theta)-\tilde{g}\}^2 ,$$

which explicitly expresses the dependence of the minimum value of (1.1) on $\underline{\underline{\tau}}$.

2 B. In the particular case of a random regression coefficient model, as defined in subsection 1 B, it is convenient to introduce the notations

$$\underline{\underline{\beta}} = E\{\underline{\underline{B}}(\Theta)\} \quad \text{and} \quad \underline{\underline{\Lambda}} = \mathrm{Var}\{\underline{\underline{B}}(\Theta)\} \tag{2.6}$$

Recalling that we want to estimate $g(\Theta) = \underline{y}'\underline{\underline{B}}(\Theta)$, we now have

$$\gamma = \underline{y}'\underline{\underline{\beta}} \;. \tag{2.7}$$

The quantities defined in (2.2) and (2.3) become

$$\underline{\underline{M}}(\Theta) = \underline{\underline{Y}}\,\underline{\underline{B}}(\Theta) \;, \qquad \underline{\underline{\mu}} = \underline{\underline{Y}}\,\underline{\underline{\beta}} \tag{2.8}$$

$$\underline{\underline{\xi}} = \underline{\underline{Y}}\,\underline{\underline{\Lambda}}\,\underline{\underline{y}} \;. \tag{2.9}$$

By (2.8) we have

$$\underline{\underline{\Gamma}} = \underline{\underline{Y}}\,\underline{\underline{\Lambda}}\,\underline{\underline{Y}}' \tag{2.10}$$

Substituting from (2.7) - (2.10) in (2.4) and (2.5), we find that the linear credibility estimator now becomes

$$\tilde{g} = \underline{y}'\underline{\underline{\beta}} + \underline{y}'\underline{\underline{\Lambda}}\,\underline{\underline{Y}}'(\underline{\underline{\Delta}}+\underline{\underline{Y\Lambda Y}}')^{-1}(\underline{\underline{X}}-\underline{\underline{Y\beta}}) \;. \tag{2.11}$$

When the order m of $\underline{\underline{X}}$ is not less than the number q of regression coefficients and $\underline{\underline{Y}}$ has full rank q , we obtain a particular simple expression. We shall need the identity

$$(\underline{\underline{I}}_m+\underline{\underline{CD}}')^{-1} = \underline{\underline{I}}_m - \underline{\underline{C}}(\underline{\underline{I}}_q+\underline{\underline{D}}'\underline{\underline{C}})^{-1}\underline{\underline{D}}' \;, \tag{2.12}$$

valid for any matrices $\underline{\underline{C}}$ and $\underline{\underline{D}}$ of order $m \times q$, in the sense that the existence of one of the inverses indicated implies the existence of the other. Consider the matrix product

$$\underline{\underline{Y}}'(\underline{\underline{\Delta}}+\underline{\underline{Y\Lambda Y}}')^{-1} = \underline{\underline{Y}}'\underline{\underline{\Delta}}^{-1}(\underline{\underline{I}}+\underline{\underline{Y\Lambda Y}}'\underline{\underline{\Delta}}^{-1})^{-1} \;,$$

which occurs in (2.11). By putting $\underline{\underline{C}} = \underline{\underline{Y}}$ and $\underline{\underline{D}}' = \underline{\underline{\Lambda}}\,\underline{\underline{Y}}'\underline{\underline{\Delta}}^{-1}$ in the parantheses on the right side and using (2.12), this expression becomes

$$\begin{aligned} &\underline{\underline{Y}}'\underline{\underline{\Delta}}^{-1}\{\underline{\underline{I}}-\underline{\underline{Y}}(\underline{\underline{I}}+\underline{\underline{\Lambda Y}}'\underline{\underline{\Delta}}^{-1}\underline{\underline{Y}})^{-1}\,\underline{\underline{\Lambda Y}}'\underline{\underline{\Delta}}^{-1}\} \\ &= \{\underline{\underline{I}}-\underline{\underline{Y}}'\underline{\underline{\Delta}}^{-1}\underline{\underline{Y}}(\underline{\underline{I}}+\underline{\underline{\Lambda Y}}'\underline{\underline{\Delta}}^{-1}\underline{\underline{Y}})^{-1}\underline{\underline{\Lambda}}\}\;\underline{\underline{Y}}'\underline{\underline{\Delta}}^{-1} \;. \end{aligned}$$

We recognize the expression in script brackets as the right side in (2.12) with $\underline{\underline{C}} = \underline{\underline{Y}}'\underline{\underline{\Delta}}^{-1}\underline{\underline{Y}}$ and $\underline{\underline{D}}' = \underline{\underline{\Lambda}}$, and hence we finally get

$$\underline{\underline{Y}}'(\underline{\underline{\Delta}}+\underline{\underline{Y}}\underline{\underline{\Lambda}}\underline{\underline{Y}}')^{-1} = (\underline{\underline{I}}+\underline{\underline{Y}}'\underline{\underline{\Delta}}^{-1}\underline{\underline{Y}}\underline{\underline{\Lambda}})^{-1}\ \underline{\underline{Y}}'\underline{\underline{\Delta}}^{-1} .$$

Substituting this in (2.11),we get

$$\begin{aligned}\tilde{g} &= \underline{y}'\underline{\beta}+\underline{y}'\underline{\underline{\Lambda}}(\underline{\underline{I}}+\underline{\underline{Y}}'\underline{\underline{\Delta}}^{-1}\underline{\underline{Y}}\underline{\underline{\Lambda}})^{-1}\ \underline{\underline{Y}}'\underline{\underline{\Delta}}^{-1}(\underline{X}-\underline{\underline{Y}}\underline{\beta}) \\ &= \underline{y}'[\underline{\beta}+\underline{\underline{\Lambda}}\{(\underline{\underline{Y}}'\underline{\underline{\Delta}}^{-1}\underline{\underline{Y}})^{-1}+\underline{\underline{\Lambda}}\}^{-1}(\underline{\underline{Y}}'\underline{\underline{\Delta}}^{-1}\underline{\underline{Y}})^{-1}\,\underline{\underline{Y}}'\underline{\underline{\Delta}}^{-1}(\underline{X}-\underline{\underline{Y}}\underline{\beta})] .\end{aligned}$$

This may be written

$$\tilde{g} = \underline{y}'\{\underline{\underline{\Psi}}\hat{\underline{B}} + (\underline{\underline{I}}-\underline{\underline{\Psi}})\underline{\beta}\} , \tag{2.13}$$

where

$$\underline{\underline{\Psi}} = \underline{\underline{\Lambda}}\{(\underline{\underline{Y}}'\underline{\underline{\Delta}}^{-1}\underline{\underline{Y}})^{-1}+\underline{\underline{\Lambda}}\}^{-1} = \underline{\underline{\Lambda}}\underline{\underline{Y}}'\underline{\underline{\Delta}}^{-1}\underline{\underline{Y}}(\underline{\underline{I}}+\underline{\underline{\Lambda}}\underline{\underline{Y}}'\underline{\underline{\Delta}}^{-1}\underline{\underline{Y}})^{-1} ,$$

and

$$\hat{\underline{B}} = (\underline{\underline{Y}}'\underline{\underline{\Delta}}^{-1}\underline{\underline{Y}})^{-1}\ \underline{\underline{Y}}'\underline{\underline{\Delta}}^{-1}\underline{X} .$$

We recognize $\hat{B}$ as the generalized least squares estimator, which minimizes the quadratic form $(\underline{X}-\underline{\underline{Y}}\underline{B})'\underline{\underline{\Delta}}^{-1}(\underline{X}-\underline{\underline{Y}}\underline{B})$. If $\underline{\underline{\Delta}} = \mathrm{Var}\ (\underline{X}|\Theta=\theta)$ for some $\theta \in \Omega$, then $\hat{\underline{B}}$ is an optimal estimator of $\underline{B}(\theta)$ in the Gauss-Markov sense when $\Theta = \theta$. Thus the estimator $\tilde{g}$ of $\underline{y}'\underline{B}(\Theta)$ is a weighed average of the least squares estimator $\underline{y}'\hat{\underline{B}}$ and the a priori estimator $\underline{y}'\underline{\beta}$.

2 C. In the simplest possible version of the experience rating problem, the total claim amounts $X_1,\ldots,X_m$ reported by a risk in the first m insurance periods are assumed to be independent and identically distributed when Θ is fixed. On basis of these observations the individual pure premium $g(\Theta) = E(X_1|\Theta)$ is to be estimated. In this case the model in subsection 2 B applies with $\underline{y}$, $\underline{B}(\Theta)$, $\underline{\underline{\Lambda}}$, $\underline{\underline{\Delta}}$ and $\underline{\underline{Y}}$ equal to 1, $g(\Theta)$, $\mathrm{var}\{E(X_1|\Theta)\}$, $E\{\mathrm{var}(X_1|\Theta)\}$ and the $m \times 1$ - vector $(1,\ldots,1)'$ respectively. It is easily verified that (2.13) in this case reduces to Bühlmann's (1970) now classical credibility formula

$$\tilde{g} = \frac{m}{m+k}\bar{X} + \frac{k}{m+k}\mu \;, \tag{2.14}$$

where $k = E\,\mathrm{var}(X_1|\Theta)/\mathrm{var}\,E(X_1|\Theta)$ and $\bar{X} = m^{-1}\sum_{i=1}^{m} X_i$.

3. Empirical linear credibility estimators

3 A. We now turn to the problem posed in subsection 1 C and assume that the prior measure U and possibly also the class $\mathcal{F}$ of conditional distributions are unknown. We have at our disposal the first n of a sequence $\underline{\underline{X}}_1, \underline{\underline{X}}_2, \ldots$ of independent observable vectors corresponding to independent replications $\Theta_1, \Theta_2, \ldots$ of Θ . These "collateral" data are assumed to be independent of the current Θ and $\underline{\underline{X}}$. Nevertheless they may be utilized in the estimation of $g(\Theta)$ to the extent that they contain information about the parameter $\underline{\underline{\tau}}$ which enter into the credibility formula $\tilde{g}$. Thus, for each n let $\hat{\underline{\underline{a}}}_n = (\hat{a}_{no}, \ldots, \hat{a}_{nm})$ be a function of $\underline{\underline{X}}_1, \ldots, \underline{\underline{X}}_n$, to be thought of as an estimator of the vector of credibility coefficients $\underline{\underline{\alpha}} = (\alpha_o, \ldots, \alpha_m)$. The corresponding sequence $\{\hat{g}_n\}$ defined by

$$\hat{g}_n(\underline{\underline{X}};\underline{\underline{X}}_1, \ldots, \underline{\underline{X}}_n) = \hat{\alpha}_{no} + \sum_{i=1}^{m} \hat{\alpha}_{ni} X_i \tag{3.1}$$

will be called an <u>empirical linear (EL-) estimator</u> of $g(\Theta)$. The class of EL-estimators includes the class of linear estimators defined by (1.2) as well as the Bühlmann-Straub type of estimators. Note that we have included no requirement of unbiasedness in our definition of EL-estimators. In this connection we remark that unbiasedness of the estimators of $\underline{\underline{\tau}}$ does not imply unbiasedness of the corresponding Bühlmann-Straub estimator, since the credibility coefficients are not linear functions of $\underline{\underline{\tau}}$. Our criterion for judging EL-estimators will be based on the overall expected squared error $E\{g(\Theta)-\hat{g}_n\}^2$, the expectation now being with respect to $\Theta, \underline{\underline{X}}$ and $\underline{\underline{X}}_1, \ldots, \underline{\underline{X}}_n$ for each n .

Keep in mind that, although not explicitly indicated, this expected value depends on $\underline{\underline{\tau}}$ and the distribution of $\hat{\underline{\underline{a}}}_n$. By the assumed independence of $(\Theta,\underline{\underline{X}})$ and the collateral data we have

$$E\{g(\Theta)-\hat{g}_n\}^2 = E\,E[\{g(\Theta)-\hat{g}_n\}^2 | \underline{\underline{X}}_1,\ldots,\underline{\underline{X}}_n]$$
$$\geq \rho(\underline{\underline{\tau}}) \; .$$

An EL-estimator $\{\hat{g}_n\}$ will be called an (asymptotical) <u>EL-credibility estimator</u> relative to the subclass $\mathcal{M}$ of $\mathcal{M}_2$ if

$$\lim E\{g(\Theta)-\hat{g}_n\}^2 = \rho(\underline{\underline{\tau}}) \qquad (3.2)$$

for all $(U,\mathcal{F}) \in \mathcal{M}$. (Unless otherwise is explicitly indicated, all limits here and later refer to $n \to \infty$.) This criterion is related to Robbins' (1955 and 1964) concept of asymptotical optimality of empirical Bayes procedures. A general sequence $\{\hat{g}_n\}$ of estimators is asymptotically optimal if the limit on the left side in (3.2) exists and equals the Bayes risk, which is the infimum of (1.1) taken over all possible $\hat{g}$. Since we consider only the restricted class of estimators on the form (3.1), we can only require that the limit on the left side in (3.2) equals $\rho(\tau)$, which is usually larger than the Bayes risk.

3 B. For each n let $\hat{\underline{\underline{\tau}}}_n$ be an estimator of $\underline{\underline{\tau}}$ based on $\underline{\underline{X}}_1,\ldots,\underline{\underline{X}}_n$. A reasonable way to construct an EL-estimator is to define $\{\hat{g}_n\}$ by

$$\hat{g}_n = \hat{\gamma}_n + \hat{\underline{\underline{\zeta}}}'_n(\underline{\underline{X}}-\hat{\underline{\underline{\mu}}}_n) \; , \qquad (3.3)$$

where $\hat{\underline{\underline{\zeta}}}_n$ is obtained from (2.4) by insertion of $(\widehat{\underline{\underline{\Delta}}+\underline{\underline{\Gamma}}})_n$ and $\hat{\underline{\underline{\xi}}}_n$ from $\hat{\underline{\underline{\tau}}}_n$ on the right side. Suppose the sequence of estimators $\{\hat{\underline{\underline{\tau}}}_n\}$ is consistent for $\underline{\underline{\tau}}$ in some class $\mathcal{M} \subset \mathcal{M}_2$, i.e.

$$\text{plim } \hat{\underline{\underline{\tau}}}_n = \underline{\underline{\tau}} \qquad \text{for all} \quad (U,\mathcal{F}) \in \mathcal{M}. \tag{3.4}$$

Then it follows that $\text{plim } \hat{g}_n = \tilde{g}$ for all $(U,\mathcal{F}) \in \mathcal{M}$, since $\tilde{g}$ is continuous in $(\underline{\underline{X}},\underline{\underline{\tau}})$ and $\text{plim}(\underline{\underline{X}},\hat{\underline{\underline{\tau}}}_n) = (\underline{\underline{X}},\underline{\underline{\tau}})$. We shall take particular interest in the situation when $\underline{\underline{\zeta}}$ is uniformly bounded, i.e.

$$\begin{aligned} &\|\underline{\underline{\zeta}}\| \leq k \quad \text{for some finite constant } k \\ &\text{and all} \quad (U,\mathcal{F}) \in \mathcal{M}. \end{aligned} \tag{3.5}$$

Then, by truncation if necessary, we can, and always will, arrange it so that

$$|\hat{\underline{\underline{\zeta}}}_{ni}| \leq k \quad \text{for all} \quad n \quad \text{and all} \quad i=1,\ldots,m$$

without disturbing the fact that $\text{plim } \hat{\underline{\underline{\zeta}}}_n = \underline{\underline{\zeta}}$ for all $(U,\mathcal{F})$ in $\mathcal{M}$. Further we shall assume that $\{\hat{\underline{\underline{\tau}}}_n\}$ satisfies the condition

$$\lim E(\hat{\gamma}_n-\gamma)^2 = \lim E\|\underline{\underline{\mu}}_n-\underline{\underline{\mu}}\|^2 = 0 \quad \text{for all } (U,\mathcal{F}) \in \mathcal{M}. \tag{3.6}$$

The restrictiveness of these conditions will be commented at the end of this subsection. First we state and prove the main result.

Theorem 3.7. If the conditions (3.4) - (3.6) are satisfied, then $\{\hat{g}_n\}$ defined by (3.3) is an EL-credibility estimator relative to $\mathcal{M}$.

Proof. Assume $(U,\mathcal{F}) \in \mathcal{M}$. We must establish the equality

$$\lim E\{g(\Theta)-\hat{\gamma}_n-\hat{\underline{\underline{\zeta}}}_n'(\underline{\underline{X}}-\hat{\underline{\underline{\mu}}}_n)\}^2 = \rho(\underline{\underline{\tau}}) . \tag{3.8}$$

By subtracting and adding $\gamma + \underline{\underline{\zeta}}_n'\underline{\underline{\mu}}$ within the script brackets, the left side in (3.8) can be rewritten as

$$
\begin{aligned}
& E\{g(\Theta)-\gamma-\hat{\underline{\underline{\zeta}}}_n'(\underline{\underline{X}}-\underline{\underline{\mu}})\}^2 \\
+ \; & 2\,E\{g(\Theta)-\gamma-\hat{\underline{\underline{\zeta}}}_n'(\underline{\underline{X}}-\underline{\underline{\mu}})\}\{\gamma-\hat{\gamma}_n+\hat{\underline{\underline{\zeta}}}_n'(\hat{\underline{\underline{\mu}}}_n-\underline{\underline{\mu}})\} \\
+ \; & E\{\gamma-\hat{\gamma}_n+\hat{\underline{\underline{\zeta}}}_n'(\hat{\underline{\underline{\mu}}}_n-\underline{\underline{\mu}})\}^2 \;.
\end{aligned}
\tag{3.9}
$$

First we look at the third term in (3.9). Putting $z_{n,i} = \hat{\mu}_{ni} - \mu_i$ for $i=1,\ldots,m$, $z_{n,m+1} = \gamma - \hat{\gamma}_n$ and $c = \max(1,k)$, we get

$$
\begin{aligned}
& E\{\gamma-\hat{\gamma}_n+\hat{\underline{\underline{\zeta}}}_n'(\hat{\underline{\underline{\mu}}}_n-\underline{\underline{\mu}})\}^2 \\
\leq \; & c\,E\{\sum_{i=1}^{m+1}|z_{ni}|\}^2 = c\sum_{i=j=1}^{m+1}E|z_{ni}z_{nj}| \\
\leq \; & c\sum_{i=j=1}^{m+1}\{E(z_{ni}^2)\}^{\frac{1}{2}}\{E(z_{nj}^2)\}^{\frac{1}{2}} \quad ,
\end{aligned}
$$

the last inequality following from that of Schwarz. By assumption (3.6) $\lim E(z_{nj}^2) = 0$ for $j=1,\ldots,m+1$, which shows that the third term in (3.9) converges to 0 as $n \to \infty$. Next we show that the first term in (3.9) converges to $\rho(\underline{\tau})$. Let us write $f_n = \{g(\Theta)-\gamma-\hat{\underline{\underline{\zeta}}}_n'(\underline{\underline{X}}-\underline{\underline{\mu}})\}^2$ and $f = \{g(\Theta)-\gamma-\underline{\underline{\zeta}}'(\underline{\underline{X}}-\underline{\underline{\mu}})\}^2$. We must prove that $\lim E(f_n) = E(f)$. Assume that this equality is not true, so that we can find an $\epsilon > 0$ and a subsequence $\{f_{n'}\} \subset \{f_n\}$ such that

$$
E(f_{n'}) \geq E(f) + \epsilon \quad \text{for all } n' . \tag{3.10}
$$

Since $\operatorname{plim}_{n'\to\infty} f_{n'} = f$, we can find a new subsequence $\{f_{n''}\} \subset \{f_{n'}\}$ such that $\lim_{n''\to\infty} f_{n''} = f$ almost everywhere. Since $\{f_{n''}\}$ is dominated by the integrable function $\{\,|g(\Theta)| + |\gamma| + c\,\Sigma_{i=1}^m|X_i-\mu_i|\}^2$, it follows by the dominated convergence theorem that $\lim_{n''\to\infty} E(f_{n''}) = E(f)$. This contradicts the inequality (3.10), and so we must have $\lim E(f_n) = E(f) = \rho(\underline{\tau})$.

Finally we note that the second term in (3.9) converges to 0 , as a consequence of the Schwarz inequality and the results obtained concerning the first and third term in (3.9). These things together imply (3.8). $\square$

3 C. This subsection gives sufficient conditions for the fulfilment of assumptions (3.5) and (3.6) and is thus concerned with the applicability of theorem 3.7.

Lemma 3.11. If $g(\theta)$ has an unbiased linear estimator, i.e.

$$g(\theta) = E(\sum_{i=1}^{m} c_i X_i \,|\Theta=\theta) = \underline{c}'\underline{M}(\theta) \tag{3.12}$$

for some constant, known vector $\underline{c} = (c_1,\dots,c_m)$, then condition (3.5) is satisfied.

Proof. Relation (3.12) is $g(\theta) = \Sigma_{i=1}^m c_i M_i(\theta)$. In this case $\operatorname{cov}\{g(\Theta),M_k(\Theta)\} = \Sigma_{i=1}^m c_i \operatorname{cov}\{M_i(\Theta),M_k(\Theta)\}$. By (2.3) we thus have $\underline{\xi} = \underline{\Gamma}\,\underline{c}$, and by (2.4)

$$\underline{\zeta} = (\underline{\Delta}+\underline{\Gamma})^{-1}\,\underline{\Gamma}\,\underline{c} \,. \tag{3.13}$$

To prove that $\underline{\zeta}$ is bounded, it suffices to prove that $\sup \underline{b}'\underline{\zeta} < \infty$ when $\underline{\zeta}$ is defined by (3.13) and supremum is taken over all $\underline{b}$ such that $\|\underline{b}\| = 1$ and all positive definite symmetric matrices $\underline{\Delta}$ and $\underline{\Gamma}$. For this purpose consider a fixed pair $(\underline{\Delta},\underline{\Gamma})$. By the symmetry of $\underline{\Gamma}^{-1}\underline{\Delta}$ there exists an orthonormal matrix $\underline{P}$ such that $\underline{P}\underline{\Gamma}^{-1}\underline{\Delta}\underline{P}' = Q = \operatorname{diag}(q_1,\dots,q_m)$, a diagonal matrix. The diagonal elements q_i are positive since $\underline{\Gamma}^{-1}\underline{\Delta}$ is positive definite. We easily derive that

$$\underline{b}'\underline{\zeta} = (\underline{P}\underline{b})'(\underline{Q}+\underline{I})^{-1}\underline{P}\underline{c} \,.$$

By the orthonormality of $\underline{P}$, $\|\underline{P}\,\underline{x}\| = \|\underline{x}\|$ for any $m \times 1$ -vector $\underline{x}$. Hence

$$\sup_{\|\underline{b}\|=1} \underline{b}'\underline{\zeta} \leq \sup_{\|\underline{b}\|=\|\underline{d}\|=1} \underline{b}'(\underline{Q}+\underline{I})^{-1}\underline{d}\|\underline{c}\| = \sup_{\|\underline{b}\|=\|\underline{d}\|=1} \sum_{i=1}^{m} \frac{b_i d_i}{1+q_i} \|\underline{c}\|$$

$$\leq \sup_{\|\underline{b}\|=\|\underline{d}\|=1} \sum_{i=1}^{m} b_i d_i \|\underline{c}\| = \|\underline{c}\| .$$

Since the bound $\|\underline{c}\|$ is independent of $\underline{\Delta}$ and $\underline{\Gamma}$, the lemma is proved. □

Corollary to lemma 3.11. Condition (3.5) is satisfied in the random regression coefficient model.

Proof. In the random regression coefficient model we have

$$g(\theta) = \underline{v}'\underline{B}(\theta) = \underline{v}'(\underline{Y}'\underline{Y})^{-1}\underline{Y}'\underline{M}(\theta) ,$$

the last equality following from the first relation in (2.8) and the fact that $\underline{Y}$ has full rank. Thus condition (3.12) is satisfied with $\underline{c} = \underline{Y}(\underline{Y}'\underline{Y})^{-1}\underline{v}$. □

Finally in this subsection we comment on condition (3.6). Assume that from each $\underline{X}_j$ we can construct an estimator γ_j^* such that

$$K = \sup \operatorname{var}(\gamma_j^*) < \infty \quad \text{and} \quad \lim n^{-1} \sum_{i=1}^{n} E\gamma_i^* = \gamma . \qquad (3.14)$$

Defining $\hat{\gamma}_n = n^{-1} \sum_{i=1}^{n} \gamma_i^*$, we then have

$$E(\hat{\gamma}_n-\gamma)^2 = \operatorname{var}(n^{-1} \sum_{i=1}^{n} \gamma_i^*) + \{E(n^{-1} \sum_{i=1}^{n} \gamma_i^*) - \gamma\}^2$$

$$\leq n^{-1}K + \{n^{-1} \sum_{i=1}^{n} E(\gamma_i^*) - \gamma\}^2 .$$

Hence $\lim E(\hat{\gamma}_n-\gamma)^2 = 0$. In the important special case when the pairs $(\Theta_j,\underline{X}_j)$ are independent replications of $(\Theta,\underline{X})$ and

$g(\theta)$ has an unbiased estimator $\gamma^*(\underline{\underline{X}})$ with finite second order moment, condition (3.14) is satisfied by putting $\gamma_i^* = \gamma(\underline{\underline{X}}_i)$. Similar remarks apply to the second part of condition (3.6).

3 D. As an alternative to the criterion based on the overall expected error we could judge an EL-estimator by

$$\bar{E}\{g(\Theta) - \hat{g}_n\}^2 , \tag{3.15}$$

where $\bar{E}$ denotes expectation with respect to $(\Theta,\underline{\underline{X}})$. Under assumption (3.4) the sequence $\{\hat{g}_n\}$ of estimators defined by (3.3) is asymptotically equivalent to the linear credibility estimator in the sense that the random variable (3.15) converges in probability to $\rho(\underline{\underline{\tau}})$ as $n \to \infty$. This is so since the expected squared error (1.1) is a continuous function of the coefficient vector $\underline{\underline{a}} = (a_o,\dots,a_m)$ defining $\hat{g}$ by (1.2) and the coefficient vector of $\hat{g}_n$ converges in probability to that of $\tilde{g}$.
The present optimality criterion is weaker than that based on (3.2). In fact, since $E\{g(\Theta)-\hat{g}_n\}^2 = E\,\bar{E}\,\{g(\Theta)-\hat{g}_n\}^2$ and $\bar{E}\{g(\Theta)-\hat{g}_n\}^2 \geq \rho(\tau)$, $\{\hat{g}_n\}$ can be EL-credibility estimator relative to $(U,\mathfrak{F})$ only if $\text{plim}\ \bar{E}\{g(\Theta)-\hat{g}_n\}^2 = \rho(\tau)$.

4. Estimation of the credibility coefficients

4 A. In this section we give an example of how to construct consistent estimators of the credibility coefficients. We consider the linear regression model defined in subsections 1 B and 2 B, by which $\underline{\underline{X}}$ is of the form

$$\underline{\underline{X}} = \underline{\underline{Y}}\,\underline{\underline{B}}(\Theta) + \underline{\underline{V}}$$

with

$$E(\underline{\underline{V}}|\Theta) = \underline{\underline{O}} .$$

The pair $(U, \mathcal{F})$ is assumed to belong to the subfamily $\mathcal{M}$ of $\mathcal{M}_2$ defined by the following conditions.

<u>The conditional covariance matrix of $\underline{\underline{V}}$, given $\Theta = \theta$, is of the form</u>

$$\mathrm{Var}(\underline{\underline{V}}|\Theta=\theta) = D(\theta)\,\underline{\underline{I}} ,$$

<u>where $D(\theta)$ is some realvalued, positive function.</u>

<u>The vector $\underline{\underline{X}}$ has finite moments up to fourth order.</u>

Under these assumptions we have

$$\underline{\underline{\Delta}} = E\{\mathrm{Var}(\underline{\underline{X}}|\Theta)\} = E\{\mathrm{Var}(\underline{\underline{V}}|\Theta)\} = \delta\,\underline{\underline{I}}, \tag{4.1}$$

where

$$\delta = E\{D(\Theta)\} ,$$

and by the **Hölder** inequality

$$E\{|B_i^p(\Theta)V_k^q|\} < \infty \quad \text{if} \quad p,q > 0 \quad \text{and} \quad p+q \leq 4 , \tag{4.2}$$

for arbitrary components $B_i(\Theta)$ of $\underline{\underline{B}}(\Theta)$ and V_k of $\underline{\underline{V}}$.

4 B. According to formula (2.11) and relation (4.1) we need consistent estimators of the parameters $\underline{\underline{\beta}}$, $\underline{\underline{\Delta}}$ and δ. At our disposal are data from a sequence of independent realizations of this regression situation. We subscript by j those quantities which belong to the j'th regression. Thus for the j'th regression Θ_j, $\underline{\underline{B}}_j = \underline{\underline{B}}(\Theta_j)$, $D_j = D(\Theta_j)$, $\underline{\underline{Y}}_j$, $\underline{\underline{X}}_j$ and $\underline{\underline{V}}_j$ are defined according to the general framework. The design matrices $\underline{\underline{Y}}_j$ need not be equal, and may even be of different dimensions $m_j \times q$. Except for this fact the regressions may be considered as identical replications of the current regression. For each j let $\underline{\underline{F}}_j$ be a matrix of order $m_j \times (m_j - q)$, the columns of which form an orthonormal basis of the orthocomplement of the linear space spanned by the columns of $\underline{\underline{Y}}_j$. In the well known manner we transform $\underline{\underline{X}}_j$ into the least squares estimator of $\underline{\underline{B}}_j$,

$$\hat{\underline{\underline{B}}}_j = (\underline{\underline{Y}}_j' \, \underline{\underline{Y}}_j)^{-1} \, \underline{\underline{Y}}'_j \, \underline{\underline{X}}_j \tag{4.3}$$

and the (m_j-q)-vector

$$\underline{\underline{G}}_j = \underline{\underline{F}}_j' \, \underline{\underline{X}}_j \, . \tag{4.4}$$

(This apparently requires that all $\underline{\underline{Y}}_j$ are of full rank q. In practice this simply means that for the purpose of estimation we consider only those regressions whose design matrices are of rank q. Our asymptotic considerations presupposes that in the long run infinitely many such regressions will occur.) Substituting $\underline{\underline{X}}_j = \underline{\underline{Y}}_j\underline{\underline{B}}_j + \underline{\underline{V}}_j$ in (4.3) and (4.4), we obtain

$$\hat{\underline{\underline{B}}}_j = \underline{\underline{B}}_j + \underline{\underline{C}}_j \, \underline{\underline{V}}_j \tag{4.5}$$

with

$$\underline{\underline{C}}_j = (\underline{\underline{Y}}_j' \, \underline{\underline{Y}}_j)^{-1} \, \underline{\underline{Y}}_j'$$

and, by orthogonality,

$$\underline{\underline{G}}_j = \underline{\underline{F}}'_j \, \underline{\underline{V}}_j \, . \tag{4.6}$$

We have $\mathrm{Var}(\underline{\underline{G}}_j \,|\Theta_j) = \underline{\underline{F}}_j' \, \mathrm{Var}(\underline{\underline{V}}_j \,|\Theta_j)\underline{\underline{F}}_j = \underline{\underline{F}}_j' \, D(\Theta_j)\underline{\underline{I}}\underline{\underline{F}}_j = D(\Theta_j)\underline{\underline{I}}$, and thus

$$\hat{D}_j = \|\underline{\underline{G}}_j\|^2/(m_j-q) \tag{4.7}$$

is an unbiased estimator of D_j. From the n first regressions we form the statistics

$$\begin{aligned}
\hat{\underline{\underline{\beta}}}_n &= n^{-1} \sum_{j=1}^{n} \hat{\underline{\underline{B}}}_j \quad , \\
\hat{\delta}_n &= n^{-1} \sum_{j=1}^{n} \hat{D}_j \quad , \\
\hat{\underline{\underline{\Lambda}}}_n &= n^{-1} \sum_{j=1}^{n} (\hat{\underline{\underline{B}}}_j - \hat{\underline{\underline{\beta}}}_n)(\hat{\underline{\underline{B}}}_j - \hat{\underline{\underline{\beta}}}_n)' - \hat{\delta}_n n^{-1} \sum_{j=1}^{n} (\underline{\underline{Y}}_j'\underline{\underline{Y}}_j)^{-1} .
\end{aligned} \tag{4.8}$$

4 C. By (4.5) it is seen that $\hat{\underline{\underline{\beta}}}_n$ is an unbiased estimator of $\underline{\underline{\beta}}$ and that

$$\mathrm{Var}(\hat{\underline{\underline{\beta}}}_n) = n^{-2} \sum_{j=1}^{n} (\underline{\underline{\Lambda}} + \underline{\underline{C}}_j \underline{\underline{\Delta}} \underline{\underline{C}}_j')$$

$$= n^{-1} \underline{\underline{\Lambda}} + n^{-2} \delta \sum_{j=1}^{n} \underline{\underline{C}}_j \underline{\underline{C}}_j' .$$

We assume the matrices $\underline{\underline{Y}}_j$ to be such that the elements of $\underline{\underline{C}}_j$ are bounded by some finite constant K independent of j. (Again this essentially means that we can find infinitely many such design matrices, which in practice is a completely non-restrictive assumption.) Then it is seen that $\lim \mathrm{Var}(\hat{\underline{\underline{\beta}}}_n) = \underline{\underline{O}}$, and by the Chebyshev inequality it follows that $\operatorname{plim} \hat{\underline{\underline{\beta}}}_n = \underline{\underline{\beta}}$.

4 D. To prove that $\hat{\delta}_n$ is a consistent estimator of δ, we assume that the sequence $\{m_j\}$ is bounded, which we easily can make it be. Dropping the subscript j, relation (4.6) can be written

$$G_i = \sum_{k=1}^{m} F_{ki} V_k \quad , \qquad i=1,\ldots,m-q ,$$

where G_i is the i'th element in $\underline{\underline{G}}$ and F_{ki} the element in the k'th row and i'th column of $\underline{\underline{F}}$. By orthonormality all F_{ki} are absolutely bounded by 1, and hence

$$E(G_i^4) = E(\Sigma\, F_{ri}F_{si}F_{ti}F_{ui}V_rV_sV_tV_u) \leq \Sigma\, E|V_rV_sV_tV_u| ,$$

the sums extending over all r,s,t and u between and including 1 and m. By repeated use of the Schwartz inequality we get

$$E|V_rV_sV_tV_u| \leq (E|V_rV_s|^2)^{\frac{1}{2}}(E|V_tV_u|^2)^{\frac{1}{2}} \leq E(V_r^4) ,$$

and thus

$$E(G_i^4) \leq m^4\, E(V_r^4) .$$

The right side is finite by (4.2). Still dropping the subscript j we have by (4.7)

$$E(\hat{D}^2) = E\{\sum_{i,k=1}^{m-q} G_i^2 G_k^2\}/(m-q)^2 \leq E(G_i^4) ,$$

the last inequality following from that of Schwarz. Thus we see that

$$\operatorname{var}(\hat{D}_j) \leq E(\hat{D}_j^2) \leq m_j^4\ E(V_r^4) . \tag{4.9}$$

As remarked earlier, $\hat{D}_j$ is a conditionally unbiased estimator of $D(\Theta_j)$. Hence $E(\hat{D}_j) = E\ D(\Theta_j) = \delta$, and so $\hat{\delta}_n$ defined by (4.8) is an unbiased estimator of δ. By (4.9) we have

$$\operatorname{var} \hat{\delta}_n \leq n^{-2} \sum_{j=1}^{n} m_j^4 E\ (V_r^4) ,$$

which tends to zero as $n \to \infty$ because $\{m_j\}$ is bounded. By the Chebyshev inequality it follows that $\operatorname{plim} \hat{\delta}_n = \delta$.

4 E. Finally we consider the estimator $\hat{\underline{\underline{\Lambda}}}_n$ defined in (4.8). Putting

$$\underline{\underline{W}}_j = \underline{\underline{C}}_j\ \underline{\underline{V}}_j ,$$

(4.5) becomes

$$\hat{\underline{\underline{B}}}_j = \underline{\underline{B}}_j + \underline{\underline{W}}_j$$

and $\hat{\underline{\underline{\beta}}}_n$ defined in (4.8) becomes

$$\hat{\underline{\underline{\beta}}}_n = \underline{\underline{B}}. + \underline{\underline{W}}. ,$$

the dot denoting simple average over the n first values of the subscript j. Thus

$$n^{-1}\sum_{j=1}^{n}(\hat{\underline{\underline{B}}}_j-\hat{\underline{\underline{B}}}_n)(\hat{\underline{\underline{B}}}_j-\hat{\underline{\underline{B}}}_n)'$$

$$= n^{-1}\sum_{j=1}^{n}(\underline{\underline{B}}_j-\underline{\underline{B}}.)(\underline{\underline{B}}_j-\underline{\underline{B}}.)'+2n^{-1}\sum_{j=1}^{n}(\underline{\underline{B}}_j-\underline{\underline{B}}.)(\underline{\underline{W}}_j-\underline{\underline{W}}.)'+n^{-1}\sum_{j=1}^{n}(\underline{\underline{W}}_j-\underline{\underline{W}}.)(\underline{\underline{W}}_j-\underline{\underline{W}}.)'$$

$$= \underline{\underline{A}}_1 + 2\underline{\underline{A}}_2 + \underline{\underline{A}}_3 \ . \tag{4.10}$$

As is well known, (and easy to prove),

$$\text{plim } \underline{\underline{A}}_1 = \underline{\underline{\Lambda}} \ . \tag{4.11}$$

We rewrite $\underline{\underline{A}}_2$ as

$$\underline{\underline{A}}_2 = n^{-1}\sum_{j=1}^{n}\underline{\underline{B}}_j\ \underline{\underline{W}}_j' - \underline{\underline{B}}.\ \underline{\underline{W}}.' \ . \tag{4.12}$$

Since the elements of the matrices $\underline{\underline{C}}_j$ are absolutely bounded by a constant, which is independent of j, so is also the elements of the matrices $\text{Var}(\underline{\underline{W}}_j) = \underline{\underline{C}}_j\ \text{Var}(\underline{\underline{V}}_j)\underline{\underline{C}}_j' = \delta\ \underline{\underline{C}}_j\underline{\underline{C}}'_j$. Hence $\lim \text{Var}(\underline{\underline{W}}.) = \lim n^{-2}\ \Sigma_{j=1}^{n}\ \text{Var}(\underline{\underline{W}}_j) = \underline{\underline{0}}$, and by the Chebyshev inequality it follows that $\text{plim } \underline{\underline{W}}. = \underline{\underline{0}}$. Since $\underline{\underline{B}}.$ converges in probability, it also follows that

$$\text{plim } \underline{\underline{B}}.\ \underline{\underline{W}}.' = \underline{\underline{0}} \ . \tag{4.13}$$

For each j we have

$$E(\underline{\underline{B}}_j\underline{\underline{W}}_j') = E\{\underline{\underline{B}}_j E(\underline{\underline{W}}_j'\,|\Theta_j)\} = \underline{\underline{0}}$$

and

$$\text{Var}(\underline{\underline{B}}_j\underline{\underline{W}}_j') = E(\underline{\underline{B}}_j\underline{\underline{W}}_j'\underline{\underline{W}}_j\underline{\underline{B}}_j') = E(\underline{\underline{B}}_j\underline{\underline{V}}_j'\underline{\underline{C}}_j'\underline{\underline{C}}_j\underline{\underline{V}}_j\underline{\underline{B}}_j') \ .$$

As an easy consequence of (4.2) and the boundedness of the matrices $\underline{\underline{C}}_j$, the elements of these covariance matrices are bounded by some constant which is independent of j. Again by the Chebyshev inequality we conclude that

$$\text{plim } n^{-1} \sum_{j=1}^{n} \underline{\underline{B}}_j \underline{\underline{W}}_j' = \underline{\underline{O}} \ . \tag{4.14}$$

From (4.12) - (4.14) it follows that

$$\text{plim } \underline{\underline{A}}_2 = \underline{\underline{O}} \ . \tag{4.15}$$

The last term in (4.10) is

$$\underline{\underline{A}}_3 = n^{-1} \sum_{j=1}^{n} \underline{\underline{W}}_j \underline{\underline{W}}_j' - \underline{\underline{W}}_{\cdot} \underline{\underline{W}}_{\cdot}' \ . \tag{4.16}$$

For each j we have

$$E(\underline{\underline{W}}_j \underline{\underline{W}}_j') = E(\underline{\underline{C}}_j \underline{\underline{V}}_j \underline{\underline{V}}_j' \underline{\underline{C}}_j') = \delta \underline{\underline{C}}\, \underline{\underline{C}}_j' \ .$$

By our assumptions it easily follows that the elements of the covariance matrices

$$\text{Var}(\underline{\underline{W}}_j \underline{\underline{W}}_j') = E(\underline{\underline{W}}_j \underline{\underline{W}}_j')^2 - \{E(\underline{\underline{W}}_j \underline{\underline{W}}_j')\}^2$$

are bounded by some constant independent of j . Hence, by the now familiar way of reasoning, we have

$$\text{plim } n^{-1} \sum_{j=1}^{n} (\underline{\underline{W}}_j \underline{\underline{W}}_j' - \delta \underline{\underline{C}}_j \underline{\underline{C}}_j') = \underline{\underline{O}} \ . \tag{4.17}$$

We have already shown that $\text{plim } \underline{\underline{W}}_{\cdot} = 0$, and so from (4.16) and (4.17) we get

$$\text{plim}(\underline{\underline{A}}_3 - n^{-1} \delta \sum_{j=1}^{n} \underline{\underline{C}}_j \underline{\underline{C}}_j') = 0 \ . \tag{4.18}$$

Combining (4.10), (4.11), (4.15) and (4.18) , we find that

$$\text{plim}\{n^{-1} \sum_{j=1}^{n} (\hat{\underline{\underline{B}}}_j - \hat{\underline{\underline{\beta}}}_n)(\hat{\underline{\underline{B}}}_j - \hat{\underline{\underline{\beta}}}_n)' - n^{-1} \delta \sum_{j=1}^{n} \underline{\underline{C}}_j \underline{\underline{C}}_j'\} = \underline{\underline{\Lambda}} \ .$$

To conclude that $\text{plim } \hat{\underline{\underline{\Lambda}}}_n = \underline{\underline{\Lambda}}$, we need only notice that $\underline{\underline{C}}_j \underline{\underline{C}}_j' = (\underline{\underline{Y}}_j' \underline{\underline{Y}}_j)^{-1}$ and that

$$\operatorname{plim}\{(\hat{\delta}_n-\delta)\, n^{-1} \sum_{j=1}^{n} (\underline{\underline{Y}}_j'\underline{\underline{Y}}_j)^{-1}\} = 0 ,$$

the last relation following from the boundedness of the matrices $\underline{\underline{C}}_j$.

4 D. We have established estimators that are consistent for $\underline{\underline{\beta}}$, $\underline{\underline{\Lambda}}$ and $\underline{\underline{\Delta}}$ whenever $(U,\mathcal{F})$ belongs to the class $\mathcal{M}$ defined in subsection 4 A. The pair $(U,\mathcal{F})$ belongs to $\mathcal{M}$ if our observations possess finite fourth order moments and, for given Θ, fit the commonly used homoscedastic regression model. The first of these assumptions represents no severe restriction on the class $\mathcal{M}$. The assumption of homoscedasticity is more restrictive, and the construction of estimators which are consistent also for more general $\underline{\underline{\Delta}}$ is proposed as a subject for further investigations.

4 E. Martz & Krutchkoff (1969) studied the special case when $\underline{\underline{V}}$ is distributed as $N(\underline{\underline{O}},\delta\underline{\underline{I}})$, with δ known and all design matrices $\underline{\underline{Y}}_j$ equal to the current $\underline{\underline{Y}}$. Under these assumptions they found an empirical Bayes estimator of $\underline{\underline{B}}(\Theta)$, that is an estimator based on $\underline{\underline{X}}_1,\ldots,\underline{\underline{X}}_n$ and $\underline{\underline{X}}$ which converges in probability to the Bayes estimator. Wind (1973) considered the more general model of this subsection. Assuming all $\underline{\underline{Y}}_j$ to be equal, he found an empirical estimator of $\underline{\underline{B}}(\Theta)$ which is asymptotically optimal in the sense that its expected loss, with loss function $(\underline{\underline{B}}-\hat{\underline{\underline{B}}})'\underline{\underline{\Delta}}(\underline{\underline{B}}-\hat{\underline{\underline{B}}})$, converges to the least possible expected loss.

4 F. Finally in this section we consider estimation of credibility coefficients in a <u>nested classification credibility model.</u> The vector of observations associated with the current Θ is now of the form

$$\underline{\underline{X}} = \begin{pmatrix} \underline{\underline{X}}_1^o \\ \vdots \\ \underline{\underline{X}}_m^o \end{pmatrix} ,$$

where $\underline{\underline{X}}_1^o,\dots,\underline{\underline{X}}_m^o$ are vectors of order $s \times 1$, which are independent and identically distributed for fixed Θ . We introduce γ^o, $\underline{\underline{M}}^o(\Theta)$, $\underline{\underline{\mu}}^o$, $\underline{\underline{\xi}}^o$, $\underline{\underline{\Delta}}^o$, and $\underline{\underline{\Gamma}}^o$ by replacing $\underline{\underline{X}}$ with $\underline{\underline{X}}_i^o$ in subsection 2 A. Then we can write

$$\underline{\underline{M}}(\Theta) = \begin{pmatrix} \underline{\underline{M}}^o(\Theta) \\ \vdots \\ \underline{\underline{M}}^o(\Theta) \end{pmatrix} , \qquad \underline{\underline{\mu}} = \begin{pmatrix} \underline{\underline{\mu}}^o \\ \vdots \\ \underline{\underline{\mu}}^o \end{pmatrix} , \qquad \underline{\underline{\xi}} = \begin{pmatrix} \underline{\underline{\xi}}^o \\ \vdots \\ \underline{\underline{\xi}}^o \end{pmatrix} ,$$

(vectors of order $ms \times 1$), and

$$\underline{\underline{\Delta}} = \begin{pmatrix} \underline{\underline{\Delta}}^o, \dots, \underline{\underline{0}} \\ \vdots \quad \ddots \quad \vdots \\ \underline{\underline{0}}, \dots, \underline{\underline{\Delta}}^o \end{pmatrix} , \qquad \underline{\underline{\Gamma}} = \begin{pmatrix} \underline{\underline{\Gamma}}^o, \dots, \underline{\underline{\Gamma}}^o \\ \vdots \quad \ddots \quad \vdots \\ \underline{\underline{\Gamma}}^o, \dots, \underline{\underline{\Gamma}}^o \end{pmatrix} ,$$

(matrices of order $ms \times ms$). If we divide $\underline{\underline{\zeta}}$ into m blocks corresponding to the m blocks in $\underline{\underline{X}}$, i.e.

$$\underline{\underline{\zeta}} = \begin{pmatrix} \underline{\underline{\zeta}}_1^o \\ \vdots \\ \underline{\underline{\zeta}}_m^o \end{pmatrix}$$

with each ζ_i^o of order $s \times 1$, we find from the relation $\underline{\underline{\xi}} = (\underline{\underline{\Delta}}+\underline{\underline{\Gamma}})\,\underline{\underline{\zeta}}$ that

$$\underline{\underline{\xi}}^o = \underline{\underline{\Delta}}^o\underline{\underline{\zeta}}_k^o + \sum_{i=1}^{m} \underline{\underline{\Gamma}}^o\underline{\underline{\zeta}}_i^o \qquad , \quad k=1,\dots,m .$$

Hence all $\underline{\underline{\zeta}}_k^o$ are equal to the same vector $\underline{\underline{\zeta}}^o$, which must be

$$\underline{\underline{\zeta}}^o = (\underline{\underline{\Delta}}^o+m\underline{\underline{\Gamma}}^o)^{-1}\underline{\underline{\xi}}^o .$$

The credibility formula (2.5) becomes

$$\begin{aligned} \tilde{g} &= \gamma + \underline{\underline{\zeta}}^{o\prime}\Big(\sum_{j=1}^{m} \underline{\underline{X}}_j^o - m\underline{\underline{\mu}}^o\Big) \\ &= \gamma + \underline{\underline{\xi}}^{o\prime}(\underline{\underline{\Gamma}}^o+m^{-1}\underline{\underline{\Delta}}^o)^{-1}(\underline{\underline{X}}_{\bullet}^o-\underline{\underline{\mu}}^o) \quad , \end{aligned}$$

where $\underline{\underline{X}}^o_{\cdot} = m^{-1} \sum_{i=1}^{m} \underline{\underline{X}}^o_i$. If in particular $g(\theta) = \underline{\underline{c}}'\underline{\underline{M}}^o(\theta)$, we have $\gamma = \underline{\underline{c}}'\underline{\underline{\mu}}^o$ and $\underline{\underline{\xi}}^o = \underline{\underline{\Gamma}}^o\underline{\underline{c}}$, and by substitution in $\tilde{g}$ we find

$$\tilde{g} = \underline{\underline{c}}'\{\underline{\underline{\Phi}}\underline{\underline{X}}^o_{\cdot} + (\underline{\underline{I}}-\underline{\underline{\Phi}})\underline{\underline{\mu}}^o\} ,$$

where
$$\Phi = \underline{\underline{\Gamma}}^o(\underline{\underline{\Gamma}}^o + m^{-1}\underline{\underline{\Delta}}^o)^{-1} .$$

Again Bühlmann's formula in subsection 2 C is obtained as a special case by putting $s = 1$ (and $\underline{\underline{c}} = 1$).

We consider the task of estimating the parameters $\underline{\underline{\mu}}^o$, $\underline{\underline{\Delta}}^o$ and $\underline{\underline{\Gamma}}^o$ from n collateral vectors $\underline{\underline{X}}_j = (\underline{\underline{X}}^{o\prime}_{j1},\ldots,\underline{\underline{X}}^{o\prime}_{jm})'$, $j=1,\ldots,n$, which can be considered as independent replications of the current $\underline{\underline{X}}$. From the j'th set of observations we obtain the mean vector

$$\underline{\underline{X}}^o_{j\cdot} = m^{-1} \sum_{i=1}^{m} \underline{\underline{X}}^o_{ji} .$$

A consistent estimator of $\underline{\underline{\mu}}^o$ is

$$\hat{\underline{\underline{\mu}}}^o_n = \underline{\underline{X}}^o_{\cdot\cdot} = n^{-1} \sum_{j=1}^{n} \underline{\underline{X}}^o_{j\cdot} .$$

For each j we form the empirical covariance matrix

$$\underline{\underline{S}}_j = (m-1)^{-1} \sum_{i=1}^{m} (\underline{\underline{X}}^o_{ji} - \underline{\underline{X}}^o_{j\cdot})(\underline{\underline{X}}^o_{ji}-\underline{\underline{X}}^o_{j\cdot})' .$$

We know that $E(\underline{\underline{S}}_j|\Theta_j) = Var(\underline{\underline{X}}^o_{ji}|\Theta_j)$, and so $E(\underline{\underline{S}}_j) = \underline{\underline{\Delta}}^o$. Hence

$$\hat{\underline{\underline{\Delta}}}^o_n = n^{-1} \sum_{j=1}^{n} \underline{\underline{S}}_j$$

is a consistent estimator of $\underline{\underline{\Delta}}^o$.

Finally, to define a consistent estimator of $\underline{\underline{\Gamma}}^o$, we write

$$\underline{\underline{S}}_n = (n-1)^{-1} \sum_{j=1}^{n} (\underline{\underline{X}}_{j\cdot}^{o} - \underline{\underline{X}}_{\cdot\cdot}^{o})(\underline{\underline{X}}_{j\cdot}^{o} - \underline{\underline{X}}_{\cdot\cdot}^{o})' .$$

$\underline{\underline{S}}_n$ is a consistent estimator of $\mathrm{Var}(\underline{\underline{X}}_{j\cdot}^{o}) = E\{\mathrm{Var}(\underline{\underline{X}}_{j\cdot}^{o}|\Theta)\} + \mathrm{Var}\{E(\underline{\underline{X}}_{j\cdot}^{o}|\Theta)\} = m^{-1}\underline{\underline{\Delta}}^{o} + \underline{\underline{\Gamma}}^{o}$. Hence

$$\hat{\underline{\underline{\Gamma}}}_n^{o} = \underline{\underline{S}}_n - m^{-1} \hat{\underline{\underline{\Delta}}}_n^{o}$$

is a consistent estimator of $\underline{\underline{\Gamma}}^{o}$. If the number of components $\underline{\underline{X}}_{ji}^{o}$ in $\underline{\underline{X}}_j$ depends on j, the above estimates should be modified in a trivial manner. Under normality assumptions Norberg (1976) has studied optimality properties of statistical procedures for such nested classification models in the linear regression case.

4 G. By reconsideration of the two foregoing estimation problems it is seen that only estimation of $\underline{\underline{\xi}}$ poses real difficulties. When $g(\theta)$ possesses an unbiased estimator $\underline{\underline{c}}'\underline{\underline{X}}$, we have $\underline{\underline{\xi}} = \underline{\underline{\Gamma}}\,\underline{\underline{c}}$, and the problem reduces to that of estimating $\underline{\underline{\Gamma}}$. A consistent estimator of $\underline{\underline{\Gamma}}$ was obtained in 4 A - 4 C by posing restrictions on $\underline{\underline{\Delta}}$ and in 4 E by assuming a nested design. In some situations, when $\mathcal{F}$ is assumed to be a parametric family of distributions, $\underline{\underline{\Gamma}}$ can be estimated without such assumptions. As an example let Θ be scalar valued and the components $X_1,\ldots,X_m$ of $\underline{\underline{X}}$ independent and exponentially distributed with density $\theta^{-1}\exp(-x/\theta)$, $x > 0$, when $\Theta = \theta$. Then $E(X_1|\Theta) = \Theta$ and $\mathrm{var}(X_1|\Theta) = \Theta^2$. According to formula (2.14) we need consistent estimators of $\mu = E(\Theta)$ and $k = E(\Theta^2)/[E(\Theta^2) - E^2(\Theta)]$. Suppose now that the collateral data are just scalars, i.e. $\underline{\underline{X}}_j$ consists just of one component X_{j1}, and assume that the pairs (Θ_j, X_{j1}) are independent replications of (Θ, X_1). Now $\hat{\mu}_n = n^{-1}\Sigma_{j=1}^{n} X_{j1}$ is a consistent estimator of μ. A consistent estimator of $\mathrm{var}(X_1) = E(\Theta^2) + \mathrm{var}(\Theta) = 2E(\Theta^2) - E^2(\Theta)$, is the empirical variance

$$S_n = (n-1)^{-1} \sum_{j=1}^{n} (X_{j1} - \hat{\mu}_n)^2,$$

and thus $\hat{k}_n = (S_n+\hat{\mu}_n^2)/(S_n^2-\hat{\mu}_n^2)$ is a consistent estimator of k. In this situation we have obtained a consistent estimator of $\underline{\underline{T}}$ without estimating $\underline{\underline{\Gamma}}$ by the aid of nested observations.

5. Some remarks on EL-estimators and credibility theory.

5 A. In a discussion of the Bühlmann-Straub procedure Taylor (1974) remarks that it lacks the methodology which the other credibility formulas possess, since "the set of admissible estimators is not described in advance, and nor is the criterion by which the optimal estimator is chosen". These objections are substantial and necessitates a clarification of the point of view underlying this paper. In the first place we note that we have in fact adopted an optimality criterion, namely that defined by (3.2). This optimality criterion is based on the total expected squared error, which for all estimators not depending on the collateral data $\underline{\underline{X}}_1,\ldots,\underline{\underline{X}}_n$ coincides with the generally accepted criterion based on (1.1). As explained in subsection 3 A, this essentially amounts to adopting the point of view of empirical Bayes theory. In the second place it should be underlined that the aim of the paper is the quite pragmatic one to investigate the properties, as measured by the criterion (3.2), of Bühlmann-Straub estimators and more general EL-estimators. The reason for undertaking this task can be found in the same discussion by Taylor as he writes: "Its (the Bühlmann-Straub procedure) pragmatic value is undoubted, and, in fact, the author himself would probably make use of it if faced with a practical experience rating problem." And indeed, what else is there to do? We note that the use of linear premium formulas of the type (1.2) in insurance arose from the need of a formula which is easily calculable and also

easily understood by the policyholder. Thus purely practical points of view determine the class of formulas from which the optimal formula should be chosen. Now efficient insurance management requires that the statistics of the risk portfolio is permanently reviewed and analyzed, this typically involving as a routine the calculation of reliable estimates of the basic risk parameters $\underline{\underline{T}}$. Thus the actuary has in fact at his disposal estimates of the optimal set of coefficients in (1.2), and he may equally well use these as any nonstochastic set of coefficients. Consequently the EL-credibility estimator joins the class of estimators satisfying the mentioned practical requirements, which explains the motive of section 3.

References.

Bühlmann, H. (1970). Mathematical methods in risk theory. Springer Verlag, Berlin.

Bühlmann, H. & Straub, E. (1970). Glaubwürdigkeit führ Schadensätze. Mitteilungen der Vereinigung Schweizerischer Versicherungsmathematiker 70, 111-133.

Hachemeister, C.A. (1975). Credibility for regression models with application to trend. Credibility: Theory and Applications. Proc. Berkeley Advanced Research Conf. Credibility. Academic Press, New York.

Jewell, W.S. (1973). Multidimensional credibility. University of Calif., Berkeley, Operations Research Center, Research Report 73-7.

Maritz, J.S. (1970). Empirical Bayes Methods. Methuen and Co. Ltd., London.

Martz, H.F. Jr. & Krutchkoff, R.G. (1969). Empirical Bayes estimators in a multiple linear regression model. Biometrika 56, 367-374.

Norberg, R. (1976). Inference in random regression coefficient models with one-way and nested classifications. Scand.Jour.Statist., (to appear).

Robbins, H. (1955). An empirical Bayes approach to statistics. Proc. Third Berkely Symposium on Math.Statistics and Probability 1, 157-163. Univ. of California Press.

Robbins, H. (1964). The empirical Bayes approach to statistical problems. Ann.Math.Statist., 35, 1-20.

Swamy, P.A.V.B. (1971). Statistical inference in random coefficient regression models. Lecture notes in operations research and mathematical systems, 55. Springer Verlag, Berlin.

Taylor, G.C. (1974). In search of a general parameterfree credibility formula. Research paper No. 61. 1974, Macquarie University, School of economics and financial studies.

Taylor, G.C. (1975). Abstract credibility. Research paper No. 77, 1975, Macquarie University, School of economics and financial studies.

Wald, A. (1947). A note on regression analysis. Ann.Math.Statist. 18, 586-589.

Wind, S. (1973). An empirical Bayes approach to multiple linear regression. Ann.Math.Statist. 1, 93-103.